General cognitive abilities in orangutans (Pongo abelii and Pongo pygmaeus)

Damerius, Laura A ; Burkart, Judith M ; van Noordwijk, Maria A ; Haun, Daniel B M ; Kosonen, Zaida K ; Galdikas, Biruté M F ; Saraswati, Yenny ; Kurniawan, Denny ; van Schaik, Carel P

DOI: https://doi.org/10.1016/j.intell.2018.10.008

General Cognitive Abilities in Orangutans (*Pongo abelii and Pongo pygmaeus*)

Laura A. Damerius[1], Judith M. Burkart[1], Maria A. van Noordwijk[1], Daniel B. M. Haun[2], Zaida K. Kosonen[3], Biruté M. F.Galdikas[4], Yenny Saraswati[5], Denny Kurniawean[6] and Carel P. van Schaik[1]

[1] *Department of Anthropology, University of Zurich, Switzerland*
[2] *Leipzig Research Center for Early Child Development, University of Leipzig, Germany*
[3] *FUB-Research Group for Environmental Monitoring, Rapperswil, Switzerland*
[4] *Department of Archaeology, Simon Fraser University, Burnaby, BC, Canada*
[5] *Sumatran Orangutan Conservation Programme, Yayasan Ekosistem Lestari, Medan, North Sumatra, Indonesia*
[6] *Borneo Orangutan Survival Foundation, Nyaru Menteng, Palangka Raya, Central Kalimantan, Indonesia*

Corresponding author: Laura A. Damerius, Department of Anthropology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; Tel.: +41 44 635 58 87, Fax.: +41 44 635 68 04, E-mail: laura.damerius@uzh.ch

For over a century, theories of human intelligence have concentrated on a single general factor, the psychometric *g*, which is used to estimate reasoning ability and cognitive flexibility, i.e. general intelligence. To better understand the evolution of general intelligence, it is important to identify the presence of a psychometric *g* in nonhuman animals, especially in primates, and to further disentangle the influences affecting its development. We therefore investigated the cognitive abilities of 53 Bornean and Sumatran orangutans to assess the presence of a psychometric *g*, and to explore possible influences on its expression. We did so using a set of carefully selected physical cognition tasks addressing abilities of inhibitory control, behavioral flexibility, causal reasoning, tool use, and associative- and reversal learning, and presented tasks to the subjects in the absence of human experimenters. A principal component analysis of the individuals' performances revealed a single component, which accounted for 31% of the individual variation in task performance. This *g* could not be explained by non-cognitive confounding variables, such as health status, island of origin, or rearing background. Furthermore, we found a modest correlation between an individual's independently assessed curiosity and *g*, which is consistent with the notion that accumulating experience affects the developmental construction of *g.* Together, our results suggest there is evidence for general intelligence in orangutans comparable to humans and chimpanzees, and thus evolutionary continuity in this trait.

Keywords: general intelligence, *g*, positive manifold, cognitive development

Introduction

General intelligence in humans can be broadly defined as the ability to show behavioral flexibility, to think abstractly, and comprehend complex ideas; it is thought to be based on reasoning, problem-solving, planning, and learning from experience (Gottfredson 1997). For over a century, theories of general intelligence relied on the reliable emergence of a single general factor, the psychometric *g*, from factor analysis of the performances across different tasks and domains (Spearman 1904, 1927, Carroll 1993, Jensen and Weng 1994, Jensen 1998).

Although the existence of *g* as a measurable phenomenon of across-domain correlations between diverse cognitive tests is widely accepted, there is still active debate about whether it represents a latent variable or rather some emerging property (cf. Burkart et al. 2017), and about how *g* is related and contributing to cognitive processes (Conway and Kovacs 2015). Carroll (2003), Geary (2005), and Horn and McArdle (2007) argue in favor of either some domain-general ability or a bundle of interacting abilities, which must be highly functionally integrated to be effective (Deary et al. 2016, Burkart et al. 2017). The latter position is consistent with effects of inbreeding or health (Penke et al. 2007, Hagenaars et al. 2016). Regardless of the outcome of these debates, the predictable presence of *g* in broad psychometric studies suggests some system-level property affecting overall cognitive performance across a broad array of tasks.

Cognitive abilities are developmentally constructed and reconstructed (Vygotsky 1978). Intelligence can therefore be shaped by socio-cultural influences (Neisser et al. 1996, Nisbett 2009, Nisbett et al. 2012, Flynn 2016) and cognitive performance by specific experiences during development (Reyes-García et al. 2016). These experience effects could also explain the positive relation between motivation, including curiosity and persistence, and general cognitive performance as suggested by Gottfried (1990): Depending on how an individual approaches its surroundings, e.g. with high curiosity or great social interest, it will experience different learning opportunities. The accumulation of these experiences can influence an individual's knowledge and was suggested as explanation for the effect of early novelty seeking in human infants on subsequent IQ development (Bornstein and Sigman 1986, Vietze and Coates 1986).

It has traditionally been assumed, usually implicitly, that general intelligence is a phylogenetically derived feature of humans, and therefore unique to our species. However, there is a rapidly expanding literature investigating domain-general cognitive ability in nonhuman animals (reviewed in Burkart et al. 2017). In nonhuman animals intelligence can be broadly defined as the ability to acquire and utilize knowledge to organize effective behavior in both familiar and novel contexts, and thus to behave flexibly (Byrne 1994, Yoerg 2001, Rumbaugh

79 and Washburn 2003). An open question in animals therefore is to what extent intelligence,
80 broadly defined, corresponds to a potential psychometrically derived *g*-factor.
81 One major source of evidence in favor of a psychometric general factor in nonhuman
82 animals is provided by interspecific patterns, where species differences in cognitive
83 performance on a wide variety of tests can largely be captured by a single variable (Deaner et al.
84 2006, Reader et al. 2011). Here, we refer to this interspecific factor as *G,* to distinguish it from
85 the intraspecific *g*. Because this *G* measure is tightly correlated with brain size (Deaner et al.
86 2007) and inhibitory control (MacLean et al. 2014; see Burkart et al. 2017), both known
87 correlates of *g* in humans (Deary et al. 2010, Meldrum et al. 2017), it probably expresses very
88 similar abilities as the intraspecific *g* measure. Additionally, an increasing number of studies
89 addresses the existence of a psychometric *g* within a variety of different taxa (reviews: Chabris
90 2007, Matzel et al. 2011, Burkart et al. 2017), including dogs (Arden and Adams 2016), mice
91 (Galsworthy et al. 2002, Locurto et al. 2003, Matzel et al. 2003, Matzel et al. 2011, Wass et al.
92 2012, Matzel et al. 2017), rats (Anderson 1993), bowerbirds (Keagy et al. 2011), New Zealand
93 Robins (Shaw et al. 2015), cotton-top tamarins (Banerjee et al. 2009), rhesus macaques
94 (Herndon et al. 1997), and chimpanzees (Herrmann et al. 2010, Hopkins et al. 2014, Woodley of
95 Menie et al. 2015).
96 Although these animal studies may appear to settle the matter, many of them can be
97 criticized. Most have methodological limitations that need to be dealt with before concluding
98 that there is evidence of *g*. There is a serious risk of a false positive result, for instance because a
99 test-battery is applied where multiple tests all basically measure similar problem-solving
100 abilities, or because studies do not control for individual variation in health, motivation, or
101 habituation (Burkart et al. 2017). Alternatively, the existence of *g* can be masked, e.g. through
102 small sample sizes, or if many tasks are included that are automatized during ontogeny
103 (experience-dependent), and thus undergo so-called secondary modularization (Burkart et al.
104 2017). Furthermore, as in human studies, we have to expect the correlations among the
105 performance scores on various tests to be modest, meaning that all animal studies are
106 statistically underpowered (it took massive meta-analyses to convince critics of human
107 intelligence studies: Carroll 1993).
108 But even reliably identifying a psychometric *g* in an animal test battery does not *per se*
109 guarantee that this *g* corresponds to intelligence broadly defined, i.e. general intelligence.
110 Rather, additional validation tests are necessary, as done for humans, and to some extent mice
111 (e.g. Matzel et al. 2006, Matzel et al. 2011).
112 The purpose of this study was to examine the presence of a psychometric *g* factor in
113 orangutans, and to link it to biological and developmental variables. Orangutans are good
114 candidates for such a study. They are Asian great apes known for advanced cognitive abilities,

such as flexible tool use (van Schaik 2004), planning (van Schaik et al. 2013) and large innovation repertoires in the wild (van Schaik et al. 2006), as well as problem-solving abilities in zoological gardens (Lethmate 1977, Lehner et al. 2011). Orangutans also show extended development and a long phase of socially mediated learning (Schuppli et al. 2016a) to acquire their geographically variable skill repertoires (van Schaik et al. 1996, van Schaik et al. 2003, Jaeggi et al. 2010, Schuppli et al. 2016b), consistent with the possibility that intelligence is developmentally constructed under the influence of a variety of inputs and experiences.

Orangutans are also an important test species because like chimpanzees, they are great apes and therefore closely related to humans. Chimpanzees have been intensively studied, and initial results appeared to be mixed. Hermann et al. (2010) did not find *g* in chimpanzees with a sample of 106 individuals over multiple social and non-social tasks. Their factor model described two distinct areas: the spatial and the physical-social area. Hopkins et al. (2014), in contrast, derived a factor representing *g* using a broad intelligence battery based on similar methods (Herrmann et al. 2007) with 99 individuals. Woodly of Menie et al. (2015) reanalyzed the same dataset and corroborated the earlier interpretations for the presence of *g* in chimpanzees. In light of the different analytical procedures of the two sets of studies, the interpretation of the presence of *g* in chimpanzees now appears to be the most parsimonious one. In fact, Beran and Hopkins (2018) report evidence for *g* based on an unrotated principal component analysis, which as in humans, is related to self-control.

However, in light of the criticism of other animal results discussed above, a conservative approach is to test other great apes first to ascertain its presence among great apes generally. The first aim of this study, therefore, was to explore whether a psychometric *g* can be identified in a large sample of orangutans, by investigating a broad array of cognitive abilities from a test battery of five different physical cognition tasks.

This study introduced several novel features. First, the tasks were presented to the orangutans in the absence of human experimenters, to prevent any bias in favor of individual's familiarity with humans which might result in a false positive result (see also Schubiger et al. 2015). Second, the tasks in this study were carefully selected to minimize overlap in their domains, and to include abilities similar to ones used in human test batteries. They included tasks addressing flexibility, inhibitory/control, causal reasoning, reversal learning and goal directed tool use. We only used one measure per task or subtask, to minimize any possible statistical dependencies. Third, we only included non-social tasks, as is typically done in human *g* studies (see also Burkart et al. 2017).

Because finding evidence for a psychometric *g* factor is necessary, but not sufficient for showing the presence of general intelligence, the second aim of this study was to conduct validation tests to evaluate such a preliminary conclusion. First, we asked whether the

orangutan *g* could be an artifact of alternative non-cognitive factors possibly generating false positive results, such as health, sex, rearing background, and island of origin of the individuals. Second, general intelligence in humans is highly responsive to early experiences available to an individual (see above). We recently found that in orangutans a curious response-and-exploration style (resulting from previous care by humans and social housing) was the sole predictor for performance on each of the problem solving tasks used in this study separately (Damerius et al. 2017). Individuals with a long history of curiosity can accumulate physical and social experiences and improve their causal understanding and learning ability (Byrne 2016, Damerius et al. 2017). Thus, curiosity might indeed channel the prospecting of novel social and physical challenges that can lead to quantitatively and qualitatively different learning opportunities. Therefore, we were interested whether an individual's approach toward novelty, possibly resulting in different individual experiences when maintained during ontogeny, affects *g*. We predict that curiosity should be positively linked to *g*.

Methods

Study Subjects and Species

The study included 40 Bornean (*Pongo pygmaeus wurmbii*) and 13 Sumatran (*Pongo abelii*) orangutans housed at rehabilitation stations in Indonesia, of which 23 were females and 30 males (see Supplementary Table S1, category Data Set: conservative). The subjects' ages at testing ranged from 3.5-17 years (Supplementary Table S1). Their ages had been estimated upon arrival by the staff of each facility based on dental eruption patterns.

In order to be able to disentangle the actual cognitive abilities from basic motivationally driven performance, we only included subjects who passed the criterion of having touched all tasks at least once. This resulted in a conservative sample size of 53 individuals with no missing data point over all five tasks. (In the Supplementary Material we present an analysis of the extended data set of 57 individuals, which also includes the four individuals that participated in only four out of the five tasks [Table S1, category Data Set: extended], and show that the results are very similar).

The individuals varied in background, although all were wild-born (see Supplementary Table S1). This heterogeneity of background experiences of the individuals allowed us to examine experiential influences on *g.* We distinguished four categories of individuals. The first category ('human') included individuals, who had stayed with humans for at least six months and were subsequently brought to the rehabilitation station. The second category ('station') included those who had arrived at very young ages (<1.5 years old) and essentially grew up in

the rehabilitation station. The third category ('wild') where those who were rescued when their habitat was lost to conversion and brought directly to the station; these individuals were already older at arrival. Finally, we also had individuals whose background was unknown ('unknown'). Further information on the classification can be found in Damerius et al. (2017).

Study Facilities and Housing

LD, ZK and a trained assistant (Andreas Wendl) collected the data between June 2012 and June 2014 in four facilities of three organizations across Sumatra and Central Kalimantan, Borneo.

The Sumatran orangutans (*Pongo abelii*) were studied at (a) the quarantine station in Batu M'Belin (QBM) of the Sumatran Orangutan Conservation Program (SOCP), Medan, North-Sumatra. 8 subjects were part of our study (2 females, 6 males). Their ages ranged from 5 to 10 years and we tested them in their home enclosures, as they were housed solitary. (b) One solitarily male and 4 socially housed females stationed at the release site Danau Alo, Bukit Tigapuluh, Jambi, Sumatra were part of our study. They were between 3 to 6 years of age. The release site belongs to SOCP and individuals are transferred here from the QBM. The Bornean orangutans (*Pongo pygmaeus*) were studied at (c) the Orangutan Care Center and Quarantine (OCCQ) in Pasir Panjang, Central Kalimantan. This facility is managed by the Orangutan Foundation International (OFI). The 26 subjects' ages ranged from 8 to 14.5 years, with 10 females and 16 males. Testing occurred in external test cages, since all individuals were socially housed in peer-groups of equal sex ranging from 2 to 6 individuals. (d) We tested 14 solitary housed individuals (7 female, 7 males) in their home enclosure at Nyaru Menteng Rescue Center, Palangka Raya, Central Kalimantan. This rehabilitation station is managed by the Borneo Orangutan Survival Foundation (BOSF). The 14 subjects were between the ages of 6 to 17 years.

In all stations the contact with humans was limited in order to prepare the orangutans for release. Exceptions were routine check-ups by veterinarians and cage cleaning and feeding by caretakers several times a day. In Sumatra, arriving dependent infants are introduced to peer groups after quarantine. Whereas in Bornean facilities, very young dependent immatures were placed in the station's nursery. There the infants received close physical care and supervision by human caretakers, that served as surrogate mothers. All individuals on both islands received food-related enrichment several times a week and had simple enrichment devices installed in their enclosures, such as tires, ropes and platforms. Depending on the age and stage of rehabilitation most individuals received regular forest walks.

222 *Physical cognitive test battery and procedure*

223 Based on standard physical cognition tasks, we developed a test battery of five different
224 tasks for measuring problem-solving abilities: the Box Task, Detour Reaching, Tube Trap Task,
225 Honey Tool Task and Reversal Learning. These five tasks were meant to cover distinct cognitive
226 abilities, namely flexibility response, inhibition, causal reasoning, tool use, and reversal learning,
227 including associative learning and memory. We focused on performance measures that best
228 reflected possible components of domain-general cognitive processes, and thus avoided
229 measures that might reflect success based on trial-and-error exploration. We aimed at applying
230 various tasks of different cognitive domains, to minimize the risk of re-sampling the same
231 cognitive domain repeatedly, which would result in a first PC that represents this specific
232 domain, instead of being consistent with domain-general abilities.

233 The experiments required no pre-training and, with the exception of the Tube Trap Task
234 and Reversal Learning requiring multiple sessions, were presented in a random order and only
235 once. Testing occurred on consecutive days and on a specially designed presentation table
236 (Supplementary Fig. S1). For every facility we used the same experimental tasks and procedures
237 with minor modifications of the presentation table (adjustment in size and fixation) to fit the
238 conditions of each facility. The individuals were able to interact on a voluntarily basis with the
239 apparatus since it was presented from the outside of the enclosure. This way, the individuals
240 were able to reach through the mesh or bars to freely explore the apparatus. No human was
241 present during testing to prevent a bias in favor of subjects familiar with humans. Whenever
242 possible, a familiar caretaker helped to prepare the set-up of the apparatuses outside the view of
243 the enclosures. The caretaker and the experimenter left the testing area once the apparatus, the
244 presentation table and the cameras had been prepared.

245 The cognitive test battery consisted of the following experiments:

246

247 *Box Task: Flexibility*

248 This task assessed the cognitive skills of flexibility. During the experiment five identical
249 wooden boxes, each with a lid and a sliding door (Supplementary Fig. S2), were presented to the
250 individual in two consecutive conditions. During the first presentation (learning phase of 4 min)
251 all five boxes were baited with a peanut and could only be opened by flipping the lid. Individuals
252 fulfilled the criterion for learning when four out of five boxes were flipped open. Immediately
253 after passing the Learning phase the individuals were presented the second condition of the
254 task, the flexibility phase of 7 min. Out of view of the subject all lids were locked whereas the
255 sliding doors were unlocked and all boxes baited again. Thus the previously learned solution had

become non-functional and the subjects needed to explore to find the new solution to receive the peanuts. The performance measure was the number of opened boxes in the flexibility phase.

Detour Reaching: Inhibitory control

The Detour Reaching experiment serves as a key predictive measure of problem-solving skills in human and nonhuman animals and is referred to as a classical inhibitory control task (Carlson and Moses 2001, Vlamings et al. 2010, Amici et al. 2012). For five minutes a transparent Makrolon box with a small opening on the right-hand side and a large opening on the left-hand side was presented to the individuals (Supplementary Fig. S3). Behind the small opening inside the box a preferred fruit/enrichment object was placed, too big to fit through the opening. The subject could not reach for the fruit/object directly, but had to make the detour through the large opening. Therefore, the individual had to inhibit grasping for the putative direct solution and control reaching around to retrieve the reward. The performance measure was the latency from first touching the apparatus (the moment the individual began to engage with the test) until exploring the box side with the large opening for the first time, thus, showing the ability of inhibitory control.

Tube Trap Task: Causal reasoning and Learning

Causal reasoning and learning ability were assessed by the Tube Trap Task, which was inspired by Visalberghi and Limongelli (1994). The experiment displayed six horizontally placed tubes, mounted on a vertical board (Supplementary Fig. S4). The transparent front of the tubes provided a slit wide enough for an orangutan to use a finger to slide food inside the tube to the open tube ends. However, an opaque trap was situated either left or right from the center. If the food was moved to the trap, the food fell into the trap and could no longer be retrieved (Supplementary Fig. S4,a). The subjects had to learn this causal relation and to slide the food in the correct direction (Supplementary Fig. S4,b). In three consecutive trials for 4 min each the board with the six tubes was presented to the subjects and they were given the possibility to explore a total of 18 randomly distributed tubes to learn to avoid the trap. The configuration of the six tubes was changed for each trial. The performance measure was the number of rewards retrieved over all three trials.

Honey Tool Task: Goal-directed tool use

The ability to use tools, as well as the ability of causal reasoning to select the correct tool were required to successfully solve this task. The apparatus consisted of a wooden box with two embedded transparent Makrolon tubes: a straight channel and a curved L-shaped channel. The

291 front was covered with a Makrolon plate to ensure stability and visibility of the baited honey in
292 the channels (Supplementary Fig. S5). For 10 min the subject was presented with the
293 opportunity to fish for honey out of the two channels. An individual could not reach the honey
294 without tools as both channels were too long for an orangutan's finger. Three ropes and three
295 long sticks, one of which was already inserted in the straight channel, were provided. Applying
296 the sticks, honey could be reached in the straight channel, but the stick could not be bent in the
297 curved channel. Whereas the rope was too short for the straight channel, it flexibly fitted into the
298 curved L-shaped channel. Each subject was presented with the whole test only once. We could
299 use an ordinal scale to measure performance, because the data fitted a Guttman scale, which
300 showed a reproducibility coefficient close to one (0.92), with 71% individuals perfectly fitting
301 the applied Guttman scale (listed from simplest to most complex). The following scores were
302 used: 1) inserting finger in either of the channel, 2) tracing the honey in the channel from
303 outside (revealing their understanding that there is honey insight), 3) inserting tools in either of
304 the channels, 4) goal-directed tool use outside (directly aiming for the honey with the correct
305 tool), 5) successfully dipping and feeding on honey with stick in straight channel, and 6)
306 successfully dipping and feeding on honey with rope in curved channel.
307

Reversal Learning: Associative and reversal learning

Reversal learning is a widely used paradigm for assessing associative learning and reversal learning abilities (reviewed in Izquierdo et al. 2016). At first the subjects learn that one possible combination of two stimuli (e.g. location, color) is linked to a reward whereas the second one is not. Once the individual has associated the first stimulus with the reward the contingency is reversed and the first stimulus no longer indicates the reward, but the second one does. In the current study the presentation board was equipped on either side with six round white and six square black doors (Supplementary Fig. S6). The doors could be swung open by turning them to the right or to the left and gave way a small space. Behind each door of one randomly determined side the space was baited with a peanut. Therefore the subject was rewarded, when opening a baited door. The task for the subject was to learn the association between food and location, which was enhanced by the different shapes and colors of the doors. If the subject opened at least five out of the first six doors at the correct location on two consecutive trials, it had fulfilled the criteria for learning. After the location of food was accurately associated, we tested the stability of the result by applying a reminder test the next day. Subsequently, the location of the reward was switched to the other location in order to test for reversed learning, where the same criterion of at least 5 out of 6 correct choices was used. Per trial the board was presented for 4 min applying 3 to 4 consecutive trials per day (in total up to 16 min/day). For each individual the trials lasted for a maximum of 4 days.

We constructed an ordinal scale to measure performance on the three consecutive components: learning, memory and reversal learning. Individuals that did not achieve learning received a performance score of 0, individuals that were able to learn but failed subsequent tests a 1, individuals that were able to learn and remember a 2, and individuals that were able to learn, remember and learn the reversal a 3.

Measurements of experiential effects

In addition to the cognitive tasks, we carried out independent tests to assess individual variation in curiosity. Each individual's level of curiosity was assessed through five different tasks, including a novel object test, human orientation tasks, reactions toward a snake predator model and reactions to both familiar and novel food. Each task lasted two minutes and was presented separately from the outside of the enclosure without any experimenter present. A principal component analysis of these measures yielded an individual's curiosity score. Damerius et al. (2017) provide a detailed description of the tasks to estimate curiosity and the PCA analysis.

Data extraction and statistical analyses

All experiments were videotaped with two Sony camcorders (HDR-CX200 high definition handycam 5.3 MP). The video footage was imported into the program Interact, version 9.7.5.0 (© Mangold International GmbH) for transformation and coding of the cognitive performance by LD. Data on exploration and novelty responses were also coded by LD, apart from some tasks used to generate the curiosity score, which were coded by a trained assistant (Anna Schöpfer). LD and the assistant independently coded 21% of the videos and came to a good inter-rater agreement (Cohen's Kappa: 0.701; N responses= 185, P <0.001).

First, we used descriptive statistics (means with standard deviation or medians with minimum and maximum values) on the five task-specific cognitive performance scores to explore the distribution of the variance in performance (Table 1). The signs of latencies of the Detour Reaching Task were reversed for the model to represent the ability of inhibitory control.

Second, we calculated Spearman-Rho correlations to estimate the relationships between the different five measures of cognitive performance from the five different tasks. We applied Bonferroni corrections on the level of 0.01 (α/N=.05/5).

Third, in order to investigate the psychometric structure we applied an unrotated principal component analysis (PCA) on the five performance measures (Table 1), for both, the conservative and extended data set (see Supplementary Table S1). The PCAs were conducted in *IBM SPSS Statistics*, version 22.0.0.0 (© IBM Cooperation and other(s), 1989, 2013). This PCA method was chosen to answer the primary question, i.e. whether there are across-domain correlations, shared variance, between the diverse performance measures and to how many components the cognitive domains can be reduced. We subsequently performed a parallel analysis in R version 2.1 (R Core Team 2016) using the fa.parallel function from the 'psych' package (Revelle 2017) in order to determine the number of significant components. We generated 1000 randomly simulated data sets with equal n and k by resampling and generating random normal data.

Additionally, to assess whether the performance variables' distributions affected the outcome of a psychometric structure of *g*, we ranked each subject's performance within any given task and performed an additional PCA (presented in the Supplementary Material). Subjects with the best performance received the value '1' whereas subjects with the lowest performance the rank of '53'. Tied performances were given equal scores, such that the median value was returned (overview of measurements see Supplementary Table S2). For additional investigations if the factorial method used showed an effect, we conducted an exploratory factor analysis (EFA) in *JMP*, version 12.1.0 (© 2015 SAS Institute Inc.) using maximum likelihood estimation (ML) on the five performance measures (Table 1).

379 Fourth, we used Linear Models (LM) to investigate whether variation in the principal
380 component (PC1) can be explained by non-cognitive factors such as health characteristics (e.g.
381 loss of hair or stereotypical behavior patterns), species (island of origin), background, sex or age.
382 The contrasts for the categorical predictor variables *Background* was specified *a priori* to
383 explicitly compare wild individuals toward all others, individuals with 'unknown' background
384 toward the 'station' and 'human' backgrounds, and the direct comparison between 'Station' and
385 'Human' backgrounds.

386 Finally, to assess additional experiential influences on the PC1, we used univariate Linear
387 Mixed-Effects Models (LMM) with curiosity and age as fixed effects, the different rehabilitation
388 stations and backgrounds as random effects, and the principal component as a response. The
389 LMMs were first performed on the complete data set (N=52, one individual hat no curiosity
390 score) and secondly on a reduced data set of only non-wild individuals N=50.

391 All linear models were performed in *R* version 2.1 (R Core Team 2016) using the 'lme4'
392 (Bates et al. 2014) and 'multcomp' packages (Hothorn et al. 2016).

393

394

395 *Ethical note*

396 The experimental protocols were in full compliance with the Swiss Animal Welfare
397 legislation and were approved by the Indonesian Ministry of Research and Technology
398 (RISTEK). They also fully complied with ethical guidelines of each study facility and followed the
399 International Primatological Society's Ethical Standards for Research on Nonhuman Primates.
400

401

402 ## *Results*

403 *Variation in cognitive performance*

404 Table 1 shows the overall level of task performance in the five tasks of the test battery.
405 There was high variance within all of the task performances and importantly neither floor nor
406 ceiling effects were present.

407

Table 1: Description of tasks and measurements

Cognitive ability tested	Task administered	Performance measurement	Measurement description	Mean (SD) or Median (min/max)	N
Flexibility	Box Task	Quantity	Number of boxes opened (max: 5)	3.02 (1.67)	53
Inhibitory Control	Detour Reaching	Latency	Latency to exploration of the non-food side (max: 300 sec.); since short latencies represent better task performance, latencies were reversed for the model to represent the ability of inhibitory control	70.86 (91.95)	53
Tool use	Honey Tool Trap	Levels of solving	Guttman scale of goal directed tool use (max: 6)	3 (0/5)	53
Learning Remembering Reversal Learning	Reversal Learning	Levels of solving	Level of achievement on learning, remembering, and reversal learning tasks (max: 3)	2 (0/3)	53
Causal Reasoning	Tube Trap Task	Quantity	Number of rewards retrieved in all three trials (max: 18)	6.698 (3.30)	53

408

409 *Relationships between the different tasks*

410 Consistent with findings of previous human studies (Carroll 1993), the majority of
411 individual task performances were positively correlated in a pairwise comparison over all tasks
412 (Supplementary Table S3). The ability of Learning, Remembering, Reversal Learning was weakly
413 correlated with the ability of Causal Reasoning (*Spearman-Rho*, r= .333, P=.015, N=53) and
414 Flexibility (*Spearman-Rho*, r= .247, P<.074, N=53). After Bonferroni correction (at a significance
415 level of p< .01 [α/N=.05/5]), none of the correlations remained significant. There was one, very
416 weak, negative correlation between Tool Use and Learning/Remembering/Reversal Learning
417 (*Spearman-Rho*, r= -.028, P=.932, N=53), indicating little cognitive overlap between these
418 abilities.

419

420 *Principal Component Analysis*

421 Across the five different cognitive tasks, the potential overlap of the cognitive
422 performances was assessed using an unrotated Principal Component Analysis (PCA). Parallel
423 analysis revealed that only the first component (PC1) could be extracted (its eigenvalue of 1.56
424 exceeded the 95th percentile of the eigenvalues obtained from randomly generated data, see

425 also Supplementary Fig. S7). This PC1 explained 31.28% of the variation in orangutan
426 performance over all 53 individuals (Table 2). All tasks loaded positively on PC1, and item
427 loadings greater .50 were considered salient. Especially the three domains of Reversal Learning
428 Task (Learning, Memory and Reversal Learning), Causal Reasoning, and Flexibility showed high
429 loadings, followed by the ability of Inhibitory Control.

430 Similar results were obtained when performing a PCA with the extended data set of 57
431 individuals (Supplementary Table S4), which included four imputed data points (Supplementary
432 Table S1, category Data Set: extended). In addition, in order to investigate whether the variable
433 distribution or the factorial method chosen affected the outcome of the *g* analysis, we performed
434 two additional analyses on the conservative dataset: an unrotated PCA with ranks applied, and
435 an exploratory factor analysis (EFA). In both analyses, the results of the first component or
436 factor are very similar in their magnitude and orientation of the item loadings (see
437 Supplementary Tables S5 and S6). The main difference was that the percentage of shared
438 variance for the first component was slightly lower in the EFA; this difference between EFA and
439 PCA is not surprising considering their different communality estimates, the low number of
440 variables (respectively for statistical methods), and low communalities (<0.4) of the sample (but
441 see Stevens 2002, Field 2013 p. 638). In conclusion, we found robust evidence for the presence
442 of a psychometric *g*, but further validation is necessary to conclude we found evidence for
443 general intelligence.

444

Table 2: Principal Component Matrix with Component Scores and Item Loadings of the five problem-solving domains

Item	Item Loadings
	g* *(PC1)
% of variance	31.28
eigenvalues	1.56
Learning & Remembering & Reversal Learning	**.755**
Causal Reasoning	**.665**
Flexibility	**.513**
Inhibitory Control	.451
Tool use	.292

N = 53, rotation = none. Principal component (PC) loadings greater than or equal to |0.5| *appear in bold*

445

446 *PC1 scores and non-cognitive factors*

447 We next tested whether alternative non-cognitive factors possibly explain the PC1
448 scores, which we interpreted as *g*. Table 3 shows the results of a Linear Model of the PC1.
449 Because the sample size did not allow us to simultaneously include all independent variables of
450 interest, such as health, we repeated the analysis with health characteristics included, instead of
451 age (Supplementary Table S7). All analyses and other combinations of replaced variables (not
452 reported here) showed similar robust results. First, there was no significant effect of age.
453 Second, individuals with visible health characteristics did not score significantly lower on the
454 PC1. Likewise, we found no influence of background conditions (wild, time with humans as pets,
455 grown up in station, unknown), no differences between the sexes, or between the two different
456 islands. However, we did find differences between the rehabilitation stations (Fig. 1; Estimate ±
457 SE= -0.451±0.18, P=0.016), specifically between the two Bornean stations. We thus noticed that
458 current management regimes did affect PC1 scores.
459

459

460

Table 3: Linear Model of PC1 *(g)*, N=53

	Estimate	***Std. Error***	***P value***
(Intercept)	-0.105	0.65	0.874
Age	0.023	0.06	0.706
Sex (male)	-0.034	0.27	0.902
Background			
Wild vs. Others	-0.059	0.18	0.740
Others vs. Unknown	-0.032	0.10	0.762
Station vs. Human	0.103	0.23	0.655
Rehabilitation Station			
Sumatra vs. Borneo	0.238	0.16	0.153
Borneo 1 vs. Borneo 2	-0.451	0.18	**0.016 ***

F-statistic: 2.263 on 7 and 45 DF, P<.05. Significant P value is in bold.

461

462

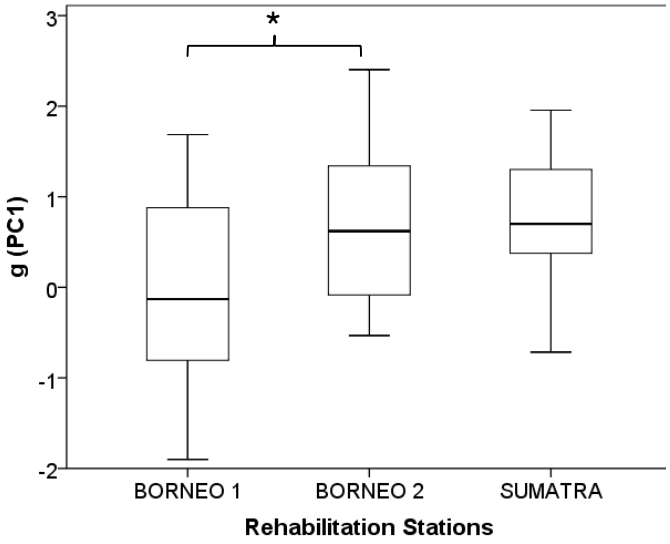


Figure 1: Box plots of the individual's *g* (PC1) over the three different rehabilitation organizations.

463

464 *Experiential effects on PC1 scores*

465 We aimed at investigating whether experiential effects have an influence on the
466 development of the possible domain-general ability captured by PC1. Therefore, we tested over
467 52 orangutans (one individual had no curiosity score) the influence of curiosity on PC1, when
468 controlling the linear mixed effect model for other potential experiential effects such as age,
469 background history and different rehabilitation station (the two latter are included as random
470 effects). Curiosity did not predict PC1 scores (Table 4).

471

471

472

Table 4: Parameter estimates and associated standard errors, obtained from a Linear Mixed-Effects Model of PC1 *(g)* over all 52 individuals

	Estimate	***SE***	***df***	***t value***	***p value***
(Intercept)	0.028	0.64	7.83	0.04	0.966
Age	0.007	0.05	18.85	0.13	0.896
Curiosity	0.193	0.18	48.50	1.07	0.292

The model is controlling for repeated observations on each rehabilitation station and background (random effects). N observations = 52, N stations = 3, N background = 4

473

474 Because wild individuals were in general strong statistical outliers, with far lower
475 curiosity (Damerius et al. 2017), and because in this study one wild individual produced a highly
476 significant outlier due to its extreme neophobia (with a curiosity score of -3, see Supplementary
477 Fig. S8), we assessed the effect of excluding the other wild individual from the analysis. Within
478 the sample of the remaining 50 non-wild individuals, an individual's curiosity positively
479 influenced PC1 scores (the *g)* in orangutans (Estimate ± SE= 0.541±0.26, P=0.0467, Table 5, Fig.
480 2). Therefore, non-wild individuals that were currently novelty-seeking and highly explorative
481 showed significant higher PC1 scores than individuals that were not (Fig. 2).

482

Table 5: Parameter estimates and associated standard errors, obtained from a Linear Mixed-Effects Model of PC1 *(g)* over 50 non-wild individuals

	Estimate	***SE***	***df***	***t value***	***p value***
(Intercept)	-0.153	0.60	5.27	-0.26	0.807
Age	0.012	0.05	9.32	0.22	0.829
Curiosity	0.541	0.26	44.57	2.05	**0.0467 ***

The model is controlling for repeated observations on each rehabilitation station (random effect). N observations = 50, N stations = 3, N background = 3
*Statistical significance is indicated as follows: *P< 0.05*

483

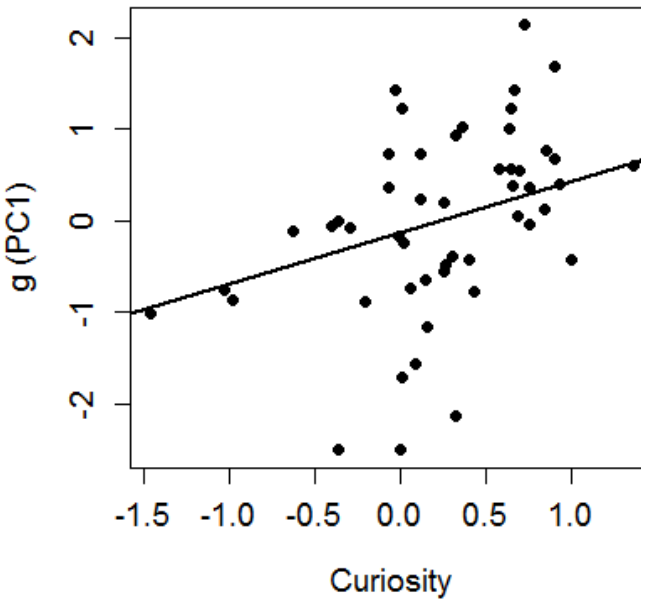


Fig. 2: The relation between curiosity and *g* (PC1) plotted with a fitted line based on the LMM (see Table 5) within all non-wild individuals (*N=50*).

Discussion

We investigated the cognitive abilities of 53 Bornean and Sumatran orangutans by comparing their performance over multiple physical cognition tasks addressing abilities of inhibitory control, behavioral flexibility, causal reasoning, tool use, and reversal learning, including associative learning and memory. Individual performance on these tasks was generally positively correlated across these tasks, and an unrotated principal component analysis indicated that the first component accounted for 31% of the individual variation in task performance of orangutans. All variables showed positive factor loadings on the first PC.

Although the five physical tasks of the test-battery were designed to measure independent abilities and minimize the transfer of experience between the tasks, one component explained a third of the variance in individual performance. The performance in associative learning, memory, reversal learning, causal reasoning, behavior flexibility and tool use all loaded positively on PC1 (Table 2). In other words, individuals performing well on one task also tended to perform well on other tasks that were designed to test distinct cognitive abilities. The absence of other significant principal components is also consistent with the notion that the performance on the various cognitive tasks reflects the effects of a single underlying or intervening variable.

Overall, the psychometric structure of these cognitive abilities in orangutans is therefore comparable to that found in humans. This conclusion is based on a small number of tests. However, we focused on highly conservative measures to minimize possible statistical dependencies. For instance, the reversal-learning test contains three phases, which have nearly independent components and provides additional information about performance. Thus, if we include associative learning as a separate measure into the principal component analysis to additionally consider the quality of learning and not only whether they learned or not, we find that the loadings remain quite similar (Supplementary Table S8). This result, taken together with the facts that a PCA based on ranked data and the EFA yielded highly similar results (Supplementary Tables S5 and S6), thus shows that the PCA results are robust. We conclude that it is justified to refer to PC1 as *g.*

Because finding the positive manifold merely constitutes a necessary, but not sufficient, condition for demonstrating the presence of general intelligence in orangutans, we also investigated whether *g* is affected by confounding variables. We did not find that possible non-cognitive variables, such as sex, health, background and origin, affected *g* and therefore consider *g* to be consistent with domain-general abilities. The one exception was an effect of the different management regimes at the stations, which we believe is of different origin, as discussed below.

So far all these analyses are concordant with the interpretation of *g* as reflecting domain-general abilities.

Relation to g studies in other animals

Our findings are generally in line with the majority of previous findings in nonhuman animals (e.g. Herndon et al. 1997, Matzel et al. 2006, Hopkins et al. 2014). They are also consistent with recent studies in chimpanzees. These studies increasingly report evidence for *g*, which, in addition and like in humans, is related to independently assessed self-control (Beran & Hopkins 2018). Intriguingly, these results are based on a very similar test battery (the PCTB, Herrmann et al. 2007) which also includes socio-cognitive tasks. This may suggest that earlier, contradicting conclusions (e.g. Herrmann et al. 2010, Herrmann and Call 2012) are the result of different analytical techniques (e.g. confirmatory factor analyses vs un-rotated PCA analyses: Herrmann et al. 2010) rather than the inclusion of socio-cognitive tasks in the battery. To allow for comparisons of studies of different species, it therefore appears crucial to standardize not only test batteries and minimal sample sizes, but also analytical techniques.

In what follows, we focus on the role of experiential factors, such as curiosity, for the construction of *g* during development and the possibility of an evolutionary continuity of domain-general processes.

Experience Effect and developmental influences on g

The ability to effectively use tools loaded only weakly on PC1 (but highly on a second PC that also had an eigenvalue > 1 but was not extracted according to the more conservative parallel analysis). A similar pattern has been reported for chimpanzees (Hopkins et al, 2014). This is consistent with the fact that tool-use abilities are prone to be automatized during ontogeny, representing so-called secondary modules that are experience-dependent (Burkart et al. 2017). To illustrate this point, consider a human example: if we were to compare bicycle riders with those who never learned to ride a bicycle, we would find no difference in average performance on a variety of tasks, but would find one for tasks that involved bicycle riding.

Previous studies suggested that curiosity underlies problem-solving performance in orangutans (Damerius et al. 2017). The results of this study show that curiosity has only a modest effect on *g* (Table 5; Figure 2), consistent with the presence of domain-general cognitive abilities. Intelligence is obviously influenced by many factors, and the modest effect indicates that not all domain-general performance is constructed based on experience due to novelty seeking and curious exploration. Damerius et al. (2017) found that the predictive effect of curiosity on problem-solving was strongest for tasks where trial-and-error exploration, rather than causal comprehension, could facilitate finding the solution. Nonetheless, when this approach is maintained during life time resulting in different opportunities for learning, this cumulative experience effect can in the long run affect an orangutan's general abilities and vice versa. Discrepancies of the *g* scores between the Bornean stations (which hosted the same sub-specie *Pongo pygmaeus wurmbii*) in all likelihood reflected these differences in opportunities for learning (Fig. 1).

Conclusion

The psychometric findings of *g* in orangutans captured the positive correlations among test scores of important problem solving components, such as the abilities of memorizing, associative learning, spatial flexibility and causal reasoning. The core definition of general intelligence captures exactly these logical problem-solving functions, reasoning- and learning abilities. In addition, the link of *g* and curiosity emphasizes the developmental construction of general-cognitive abilities. The experiences during development biased the outcome of across task performance in orangutans, as it is expected for general intelligence in humans. Thus, *g* in orangutans is in general consistent with the content of general intelligence. However, as in humans, it has to be further investigated what the emergent property of *g* exactly represents.

574 We conclude that both the content and psychometric structure of orangutan cognitive
575 abilities are comparable to those of humans. The evidence for *g* in orangutans allows us to
576 assume an evolutionary continuity in domain-general processes shared between humans and
577 great apes. Similar processes may have shaped general cognitive abilities and intelligence in
578 primates over millions of years.
579
580

581 ## *Declaration of interests*

582 The authors declare no competing financial or any other conflict of interests.

Acknowledgements

We acknowledge the collaboration and permission to collect data at the Indonesian rehabilitation stations and therefore thank the institutions of Nyaru Menteng (BOS), Orangutan Care Center and Quarantine (OCCQ), Sumatran Orangutan Conservation Program (SOCP) – Quarantine Batu M'Belin and Danau Alo Release site. In particular, we thank the station managers, animal caretakers, veterinarians and persons involved in the study at each facility, especially Ian Singleton, Matthew Nowak, Peter Pratje and Simon Husson. We thank Katja Liebal for her contribution to developing the test battery and planning of the research project. Further we thank Andreas Wendl for assisting in data collection at Nyaru Menteng and Anna Schöpfer for assisting in video coding. We also thank Erik Willems and Sereina Graber for their statistical advice. Additionally we gratefully thank our Indonesian counterpart Universitas Nasional (UNAS), especially Tatang Mitra Setia and Sri Suci Utami Atmoko and Indonesian institutions, such as the State Ministry of Research and Technology (RISTEK), and MAWAS, for their collaboration and work permissions. We are grateful to the Swiss National Science Foundation (grant 310030B_160363/1), the A.H. Schultz-Stiftung and the University of Zurich for financial support.

References

Amici, F., B. Barney, V. E. Johnson, J. Call, and F. Aureli. 2012. A modular mind? A test using individual data from seven primate species. PLoS ONE **7**:e51918.

Anderson, B. 1993. Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? Neuroscience letters **153**:98-102.

Arden, R. and M. J. Adams. 2016. A general intelligence factor in dogs. Intelligence **55**:79-85.

Banerjee, K., C. F. Chabris, V. E. Johnson, J. J. Lee, F. Tsao, and M. D. Hauser. 2009. General intelligence in another primate: individual differences across cognitive task performance in a New World monkey (Saguinus oedipus). PLoS ONE **4**:e5883.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Beran, M. J. and W. D. Hopkins. 2018. Self-control in chimpanzees relates to general intelligence. Current Biology **28**:574-579. e573.

Bornstein, M. H. and M. D. Sigman. 1986. Continuity in mental development from infancy. Child Development:251-274.

Burkart, J. M., M. N. Schubiger, and C. P. Van Schaik. 2017. The evolution of general intelligence. Behavioral and Brain Sciences:1-65.

622 Byrne, R. 1994. The evolution of intelligence. In: Behaviour and Evolution, eds. P. J.
623 B.Slater & T. R. Halliday. Cambridge University Press.
624 Byrne, R. W. 2016. Evolving Insight: How It Is We Can Think about Why Things
625 Happen. Oxford University Press.
626 Carlson, S. M. and L. J. Moses. 2001. Individual Differences in Inhibitory Control and
627 Children's Theory of Mind. Child Development **72**:1032-1053.
628 Carroll, J. B. 1993. Human cognitive abilities: A survey of factoranalytic
629 studies. New York, NY: Cambridge University Press.
630 Carroll, J. B. 2003. The higher-stratum structure of cognitive abilities: Current
631 evidence supports g and about ten broad factors. The scientific study of
632 general intelligence: Tribute to Arthur R. Jensen:5-21.
633 Chabris, C. F. 2007. 19 Cognitive and neurobiological mechanisms of the Law of
634 General Intelligence. Integrating the mind: Domain general versus domain
635 specific processes in higher cognition:449.
636 Conway, A. R. and K. Kovacs. 2015. New and emerging models of human
637 intelligence. Wiley Interdisciplinary Reviews: Cognitive Science **6**:419-426.
638 Damerius, L. A., S. Graber, E. Willems, and C. P. Van Schaik. 2017. Curiosity boosts
639 orang-utan problem-solving ability. Animal Behaviour **134**:57-70.
640 Deaner, R. O., K. Isler, J. Burkart, and C. Van Schaik. 2007. Overall brain size, and
641 not encephalization quotient, best predicts cognitive ability across non-human
642 primates. Brain, behavior and evolution **70**:115-124.
643 Deaner, R. O., C. P. Van Schaik, and V. Johnson. 2006. Do some taxa have better
644 domain-general cognition than others? A meta-analysis of nonhuman primate
645 studies. Evolutionary Psychology **4**:147470490600400114.
646 Deary, I. J., S. R. Cox, and S. J. Ritchie. 2016. Getting Spearman off the Skyhook:
647 One More in a Century (Since Thomson, 1916) of Attempts to Vanquish g.
648 Psychological Inquiry **27**:192-199.
649 Deary, I. J., L. Penke, and W. Johnson. 2010. The neuroscience of human
650 intelligence differences. Nature reviews neuroscience **11**:201.
651 Field, A. 2013. Discovering statistics using IBM SPSS statistics. Sage.
652 Flynn, J. R. 2016. Does Your Family Make You Smarter?: Nature, Nurture, and
653 Human Autonomy. Cambridge University Press.
654 Galsworthy, M., J. Paya-Cano, S. Monleon, and R. Plomin. 2002. Evidence for
655 general cognitive ability (g) in heterogeneous stock mice and an analysis of
656 potential confounds. Genes, Brain and Behavior **1**:88-95.
657 Geary, D. C. 2005. The origin of mind: Evolution of brain, cognition, and general
658 intelligence. American Psychological Association.
659 Gottfredson, L. S. 1997. Mainstream science on intelligence: An editorial with 52
660 signatories, history, and bibliography. Intelligence **24**:13-23.
661 Gottfried, A. E. 1990. Academic intrinsic motivation in young elementary school
662 children. Journal of Educational psychology **82**:525.
663 Hagenaars, S. P., S. E. Harris, G. Davies, W. D. Hill, D. C. M. Liewald, S. J. Ritchie,
664 R. E. Marioni, C. Fawns-Ritchie, B. Cullen, R. Malik, I. C. f. B. P. G.
665 Metastroke Consortium, C. SpiroMeta, C. C. A. Charge Consortium Pulmonary
666 Group, G. Longevity, B. B. Worrall, C. L. M. Sudlow, J. M. Wardlaw, J.
667 Gallacher, J. Pell, A. M. McIntosh, D. J. Smith, C. R. Gale, and I. J. Deary.
668 2016. Shared genetic aetiology between cognitive functions and physical and
669 mental health in UK Biobank (N=112[thinsp]151) and 24 GWAS consortia. Mol
670 Psychiatry **21**:1624-1632.

671 Herndon, J. G., M. B. Moss, D. L. Rosene, and R. J. Killiany. 1997. Patterns of
672 cognitive decline in aged rhesus monkeys. Behavioural brain research **87**:25-
673 34.
674 Herrmann, E. and J. Call. 2012. Are there geniuses among the apes? Philosophical
675 Transactions of Royal Society of London B: Biological Sciences **367**:2753-
676 2761.
677 Herrmann, E., J. Call, M. V. Hernàndez-Lloreda, B. Hare, and M. Tomasello. 2007.
678 Humans Have Evolved Specialized Skills of Social Cognition: The Cultural
679 Intelligence Hypothesis. Science **317**:1360-1366.
680 Herrmann, E., M. V. Hernández-Lloreda, J. Call, B. Hare, and M. Tomasello. 2010.
681 The structure of individual differences in the cognitive abilities of children and
682 chimpanzees. Psychological Science **21**:102-110.
683 Hopkins, W. D., J. L. Russell, and J. Schaeffer. 2014. Chimpanzee intelligence is
684 heritable. Current Biology **24**:1649-1652.
685 Horn, J. L. and J. J. McArdle. 2007. Understanding human intelligence since
686 Spearman. Mahwah, NJ, Lawrence Erlbaum Associates.
687 Hothorn, T., F. Bretz, P. Westfall, R. M. Heiberger, A. Schuetzenmeister, and S.
688 Scheibe. 2016. multcomp: Simultaneous Inference in General Parametric
689 Models. R package version 1.4-6.
690 Izquierdo, A., J. L. Brigman, A. K. Radke, P. H. Rudebeck, and A. Holmes. 2016. The
691 neural basis of reversal learning: An updated perspective. Neuroscience.
692 Jaeggi, A. V., L. P. Dunkel, M. A. Van Noordwijk, S. A. Wich, A. A. Sura, and C. P.
693 Van Schaik. 2010. Social learning of diet and foraging skills by wild immature
694 Bornean orangutans: implications for culture. American journal of primatology
695 **72**:62-71.
696 Jensen, A. R. 1998. The g factor: The science of mental ability. New York: Praeger.
697 Jensen, A. R. and L.-J. Weng. 1994. What is a good g? Intelligence **18**:231-258.
698 Keagy, J., J.-F. Savard, and G. Borgia. 2011. Complex relationship between multiple
699 measures of cognitive ability and male mating success in satin bowerbirds,
700 Ptilonorhynchus violaceus. Animal Behaviour **81**:1063-1070.
701 Lehner, S. R., J. M. Burkart, and C. P. van Schaik. 2011. Can captive orangutans
702 (Pongo pygmaeus abelii) be coaxed into cumulative build-up of techniques?
703 Journal of Comparative Psychology **125**:446.
704 Lethmate, J. 1977. Problem-solving behaviour in orangutans (Ponto pygmaeus).
705 Fortschritte der Verhaltensforschung.
706 Locurto, C., E. Fortin, and R. Sullivan. 2003. The structure of individual differences in
707 heterogeneous stock mice across problem types and motivational systems.
708 Genes, Brain and Behavior **2**:40-55.
709 MacLean, E. L., B. Hare, C. L. Nunn, E. Addessi, F. Amici, R. C. Anderson, F. Aureli,
710 J. M. Baker, A. E. Bania, and A. M. Barnard. 2014. The evolution of self-
711 control. Proceedings of the National Academy of Sciences **111**:E2140-E2148.
712 Matzel, L. D., Y. R. Han, H. Grossman, M. S. Karnik, D. Patel, N. Scott, S. M. Specht,
713 and C. C. Gandhi. 2003. Individual differences in the expression of a “general”
714 learning ability in mice. The Journal of neuroscience **23**:6423-6433.
715 Matzel, L. D., S. Kolata, K. Light, and B. Sauce. 2017. The tendency for social
716 submission predicts superior cognitive performance in previously isolated male
717 mice. Behavioural Processes **134**:12-21.
718 Matzel, L. D., D. A. Townsend, H. Grossman, Y. R. Han, G. Hale, M. Zappulla, K.
719 Light, and S. Kolata. 2006. Exploration in outbred mice covaries with general

learning abilities irrespective of stress reactivity, emotionality, and physical attributes. Neurobiology of Learning and Memory **86**:228-240.
Matzel, L. D., C. Wass, and S. Kolata. 2011. Individual Differences in Animal Intelligence: Learning, Reasoning, Selective Attentionand Inter-Species Conservation of a Cognitive Trait. International Journal of Comparative Psychology **24**.
Meldrum, R. C., M. A. Petkovsek, B. B. Boutwell, and J. T. Young. 2017. Reassessing the relationship between general intelligence and self-control in childhood. Intelligence **60**:1-9.
Neisser, U., G. Boodoo, T. J. Bouchard Jr, A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, and R. J. Sternberg. 1996. Intelligence: knowns and unknowns. American Psychologist **51**:77.
Nisbett, R. E. 2009. Intelligence and how to get it: Why schools and cultures count. WW Norton & Company.
Nisbett, R. E., J. Aronson, C. Blair, W. Dickens, J. Flynn, D. F. Halpern, and E. Turkheimer. 2012. Intelligence: new findings and theoretical developments. American Psychologist **67**:130.
Penke, L., J. J. Denissen, and G. F. Miller. 2007. The evolutionary genetics of personality. European Journal of Personality **21**:549-587.
R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. **URL https://www.R-project.org/**.
Reader, S. M., Y. Hager, and K. N. Laland. 2011. The evolution of primate general and cultural intelligence. Philosophical Transactions of the Royal Society of London B: Biological Sciences **366**:1017-1027.
Revelle, W. 2017. psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 1.7.3.21

Reyes-García, V., A. Pyhälä, I. Díaz-Reviriego, R. Duda, Á. Fernández-Llamazares, S. Gallois, M. Guèze, and L. Napitupulu. 2016. Schooling, Local Knowledge and Working Memory: A Study among Three Contemporary Hunter-Gatherer Societies. PLoS ONE **11**:e0145265.
Rumbaugh, D. M. and D. A. Washburn. 2003. Intelligence of Apes and Other Rational Beings. New Haven: Yale University Press.
Schubiger, M. N., F. L. Wüstholz, A. Wunder, and J. M. Burkart. 2015. High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (Callithrix jacchus). Animal cognition **18**:701-712.
Schuppli, C., S. I. Forss, E. J. Meulman, N. Zweifel, E. Rukmana, K. C. Lee, E. R. Vogel, M. A. van Noordwijk, and C. P. van Schaik. 2016a. Development of foraging skills in two orangutan populations: needing to learn or needing to grow? Frontiers in Zoology **13**:43.
Schuppli, C., E. J. M. Meulman, S. I. F. Forss, F. Aprilinayati, M. A. van Noordwijk, and C. P. van Schaik. 2016b. Observational social learning and socially induced practice of routine skills in immature wild orang-utans. Animal Behaviour **119**:87-98.
Shaw, R. C., N. J. Boogert, N. S. Clayton, and K. C. Burns. 2015. Wild psychometrics: evidence for 'general'cognitive performance in wild New Zealand robins, Petroica longipes. Animal Behaviour **109**:101-111.

769 Spearman, C. 1904. " General Intelligence," objectively determined and measured.
770 The American Journal of Psychology **15**:201-292.
771 Spearman, C. 1927. The abilities of man: their nature and measurement. Oxford,
772 England: Macmillan. **xxii**:415pp.
773 Stevens, J. 2002. Applied multivariate statistics for the social sciences. Lawrence
774 Erlbaum. Mahwah, NJ:510-511.
775 van Schaik, C. P. 2004. Among orangutans. Belknap Press of Harvard University
776 Press.
777 van Schaik, C. P., M. Ancrenaz, G. Borgen, B. Galdikas, C. D. Knott, I. Singleton, A.
778 Suzuki, S. S. Utami, and M. Merrill. 2003. Orangutan cultures and the
779 evolution of material culture. Science **299**:102-105.
780 van Schaik, C. P., L. Damerius, and K. Isler. 2013. Wild orangutan males plan and
781 communicate their travel direction one day in advance. PLoS ONE **8**:e74896.
782 van Schaik, C. P., E. A. Fox, and A. F. Sitompul. 1996. Manufacture and use of tools
783 in wild Sumatran orangutans. Naturwissenschaften **83**:186-188.
784 van Schaik, C. P., M. A. van Noordwijk, and S. A. Wich. 2006. Innovation in wild
785 Bornean orangutans (Pongo pygmaeus wurmbii). Behaviour **143**:839-876.
786 Vietze, P. M. and D. L. Coates. 1986. Information-Processing Approaches to Early
787 Identification of Mental Retardation. Annals of the New York Academy of
788 Sciences **477**:266-276.
789 Visalberghi, E. and L. Limongelli. 1994. Lack of comprehension of cause€ ffect
790 relations in tool-using capuchin monkeys (Cebus apella). Journal of
791 Comparative Psychology **108**:15.
792 Vlamings, P. J. M., B. Hare, and J. Call. 2010. Reaching around barriers: the
793 performance of the great apes and 3–5-year-old children. Animal cognition
794 **13**:273-285.
795 Vygotsky, L. 1978. Internalization of higher psychological functions. Mind in society:
796 The development of higher psychological processes:52-57.
797 Wass, C., A. Denman-Brice, C. Rios, K. R. Light, S. Kolata, A. M. Smith, and L. D.
798 Matzel. 2012. Covariation of learning and “reasoning” abilities in mice:
799 Evolutionary conservation of the operations of intelligence. Journal of
800 Experimental Psychology: Animal Behavior Processes **38**:109.
801 Woodley of Menie, M. A., H. B. Fernandes, and W. D. Hopkins. 2015. The more g-
802 loaded, the more heritable, evolvable, and phenotypically variable: Homology
803 with humans in chimpanzee cognitive abilities. Intelligence **50**:159-163.
804 Yoerg, S. I. 2001. Clever as a Fox. Animal Intelligence and what it can teach us about
805 ourselves. New York: Bloomsbury.
806
807

SUPPLEMENTARY MATERIAL

Supplementary Figures please see p. 32

Supplementary Tables

Supplementary Table S1: study subjects

Individual	Sex	Age (years)	Species	Background Category	Data set
Amin	male	6	*Pongo abelii*	unknown	conservative
Andalas	female	3	*Pongo abelii*	unknown	conservative
Ari	male	10	*Pongo pygmaeus*	human	conservative
Ayu	female	6	*Pongo abelii*	unknown	conservative
Bahruni	male	10	*Pongo abelii*	wild	conservative
Bambang	male	10.5	*Pongo pygmaeus*	human	conservative
Bella	female	14.5	*Pongo pygmaeus*	unknown	conservative
Cantik	female	10	*Pongo pygmaeus*	unknown	conservative
Ceky Chan	male	6	*Pongo abelii*	human	conservative
Cici	female	15	*Pongo pygmaeus*	unknown	conservative
Dewa	male	12	*Pongo pygmaeus*	unknown	conservative
Dora	female	3.5	*Pongo abelii*	human	conservative
Duanne	male	10	*Pongo pygmaeus*	unknown	conservative
Edwin	male	11.5	*Pongo pygmaeus*	unknown	conservative
Embrie	female	10	*Pongo pygmaeus*	unknown	conservative
Franky	male	7	*Pongo abelii*	unknown	conservative
Friend	male	5.5	*Pongo abelii*	unknown	conservative
Galih	male	10	*Pongo pygmaeus*	human	conservative
Harry	male	8.5	*Pongo pygmaeus*	station	conservative
Hulu	male	14	*Pongo pygmaeus*	human	conservative
Imas	female	9.5	*Pongo pygmaeus*	human	conservative
Jack2	male	9	*Pongo pygmaeus*	human	conservative
Jacky	female	5	*Pongo abelii*	human	conservative
Janu	male	6	*Pongo pygmaeus*	unknown	conservative
Jarot	male	5	*Pongo abelii*	station	conservative
Jill	male	8	*Pongo pygmaeus*	station	conservative
Julius	male	5	*Pongo abelii*	human	conservative
Karan	female	10	*Pongo pygmaeus*	unknown	conservative
Karen	female	15	*Pongo pygmaeus*	unknown	conservative
King	male	17	*Pongo pygmaeus*	unknown	conservative
Kraba	female	12	*Pongo pygmaeus*	unknown	conservative
Lanang	male	11	*Pongo pygmaeus*	station	conservative
Mawoto	male	17	*Pongo pygmaeus*	unknown	conservative
Mercedes	female	12	*Pongo pygmaeus*	unknown	conservative
Natalie	female	11	*Pongo pygmaeus*	unknown	conservative
Niken	female	16	*Pongo pygmaeus*	unknown	conservative
Oracle	male	8.5	*Pongo pygmaeus*	station	conservative

Otong	male	13	*Pongo pygmaeus*	unknown	conservative
Pilar	female	11	*Pongo pygmaeus*	human	conservative
Roma	female	17	*Pongo pygmaeus*	unknown	conservative
Rowland	male	10	*Pongo pygmaeus*	unknown	conservative
Runtu	female	13	*Pongo pygmaeus*	human	conservative
Sabin	male	9.5	*Pongo pygmaeus*	station	conservative
Sarimin	male	11	*Pongo pygmaeus*	unknown	conservative
Sule	female	7	*Pongo pygmaeus*	unknown	conservative
Suri	female	5	*Pongo abelii*	human	conservative
Trio	male	16	*Pongo pygmaeus*	unknown	conservative
Ulin	female	14	*Pongo pygmaeus*	unknown	conservative
Victor	male	10.5	*Pongo pygmaeus*	human	conservative
Willy	female	6	*Pongo abelii*	human	conservative
Winda	female	12	*Pongo pygmaeus*	wild	conservative
Yogi	male	8	*Pongo pygmaeus*	station	conservative
Zatarra	male	10	*Pongo pygmaeus*	unknown	conservative
Gagak	male	9.5	*Pongo pygmaeus*	station	extended (imputed Box Task)
Kasmin	female	10.5	*Pongo pygmaeus*	human	extended (imputed Box Task)
Natalia	female	18	*Pongo abelii*	wild	extended (imputed Box Task)
Inou	male	17	*Pongo pygmaeus*	unknown	extended (imputed Detour Reaching)

814

815 *Ranked dataset with 53 individuals*

Supplementary Table S2: Description of tasks and measurements with ranks applied

Cognitive ability	Task	Measurement	Median (min/max)	N
Rank Flexibility	Box Task	Number of boxes opened; largest rank describes best performance (5 boxes) and smallest worst performance (0 boxes)	27.5 (2.5/45.5)	53
Rank Inhibitory Control	Detour Reaching	Latency to first explore non-food side (show inhibit behavior); largest rank describes best performance (small latency) and smallest worst performance (no inhibition)	27 (3/51)	53
Rank Tool use	Honey Tool Trap	Guttman scale of goal directed tool use; largest rank describes best performance (successfully fishing honey) and smallest worst performance (random exploration)	30 (3.5/48.5)	53
Rank Learning Remembering Reversal Learning	Reversal Learning	Scale of learned, remembered, reversal learned; largest rank describes best performance (successfully reversal learning) and smallest worst performance (no learning, therefor no memory and no reversal learning)	25 (4/42.5)	53
Rank Causal Reasoning	Tube Trap Task	Number rewards retrieved in all three trials; largest rank describes best performance (highest number of rewards) and smallest worst performance (no reward obtained)	27.5 (1.5/52.5)	53

816

817 *Conservative data set with 53 individuals*

Supplementary Table S3:
Spearman-Rho correlation matrix of the five cognitive domains of the problem-solving tasks (*N=53*)

	Flexibility	**Inhibitory Control**	**Tool use**	**Learning Remembering Reversal Learning**	**Causal Reasoning**
Flexibility	-	.027	.099	.247	.060
Inhibitory Control		-	.080	.116	.076
Tool use			-	-.028	.114
Learning Remembering Reversal Learning				-	.333

Bonferroni niveau of significance p< .01 (α/N=.05/5)

818

819 *Extended data set with 57 individuals*

Supplementary Table S4: Principal Component Matrix of the five problem-solving domains, N=57 (imputed data set)

Item	Loadings	
	g	***Ability of Tool use***
% of variance	32.12	22.58
Eigenvalues	1.61	1.13
Learning & Remembering & Reversal Learning	**.743**	-.412
Causal Reasoning	**.679**	-.201
Flexibility	**.500**	-.069
Inhibitory Control	**.525**	**.505**
Tool use	.261	**.812**

N = 53, Rotation = none. Component loadings greater than or equal to |0.5| *appear in bold.*

820

821 *Ranked dataset with 53 individuals*

Supplementary Table S5: Principal Component Matrix of the five problem-solving domains on the ranked dataset

Item	Item Loadings	
	g	***Ability of Tool use***
% of variance	30.14	20.73
Eigenvalues	1.51	1.04
Rank Learning & Remembering & Reversal Learning	**.765**	-.376
Rank Causal Reasoning	**.676**	-.065
Rank Flexibility	**.524**	-.079
Rank Inhibitory Control	.346	.438
Rank Tool use	.267	**.832**

Note: N = 53, Rotation = none. Component loadings greater than or equal to |0.5| *appear in bold.*

822

823 *Explorative Factor Analysis with 53 individuals*

Supplementary Table S6 Exploratory Factor Analysis (Maximum-Likelihood) of the five problem-solving domains

Item	Factor Loadings
	g
% of variance	22.51
Eigenvalues	1.12
Learning & Remembering & Reversal Learning	**.687**
Causal Reasoning	**.584**
Flexibility	.407
Inhibitory Control	.333
Tool use	.191

Note: N = 53, Rotation = none. Factor loadings greater than or equal to |0.5| *appear in bold.*

824

825

826 *Conservative data set with 53 individuals*

Supplementary Table S7 (Health instead of Age): Linear Model of variable *g* (N=53)

	Estimate	***Std. Error***	***Pr(>|z|)***
(Intercept)	0.124	0.27	0.645
Health (yes)	-0.029	0.42	0.946
Sex (male)	-0.033	0.27	0.904
Background			
Wild vs. others	-0.046	0.17	0.790
others vs. Unknown	-0.045	0.10	0.663
Station vs. Human	0.088	0.23	0.706
Rehabilitation Station			
Sumatra vs. Borneo	0.191	0.10	0.074
Borneo 1 vs. Borneo 2	-0.476	0.17	**0.007****

F-statistic: 2.236 on 7 and 45 DF, p-value <.05

827

828 *Dataset with 53 individuals and LEARNING as a separate variable*

Supplementary Table S8: Principal Component Matrix of the five problem-solving domains with Learning included as a separate variable

Item	Loadings	
	g	***Ability of Tool use***
% of variance	34.39	18.38
Eigenvalues	2.06	1.10
Learning & Remembering & Reversal Learning	**.872**	-.226
Learning	**.831**	-.225
Causal Reasoning	**.581**	.062
Flexibility	.409	.064
Inhibitory Control	.258	**.654**
Tool use	.203	**.752**

Note: N = 53, Rotation = none. Component loadings greater than or equal to |0.5| *appear in bold.*

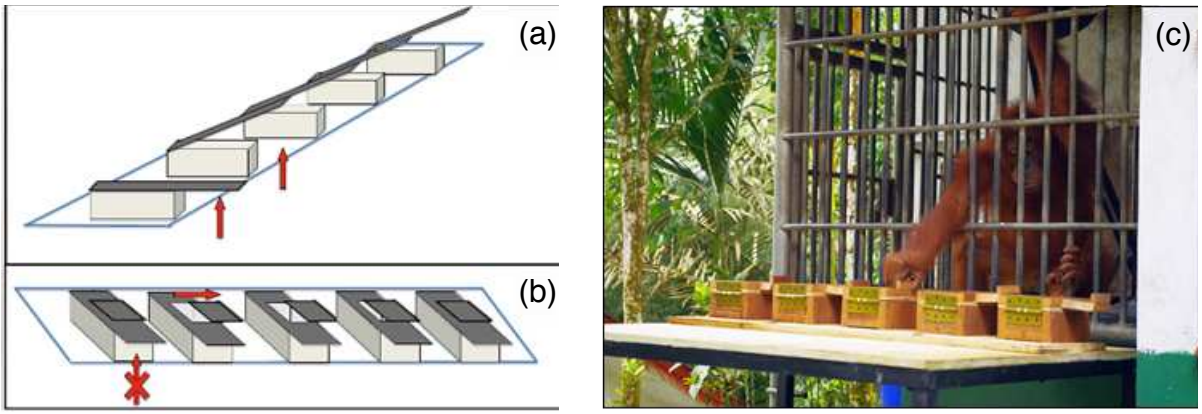
829

830 *Supplementary Figures*

831

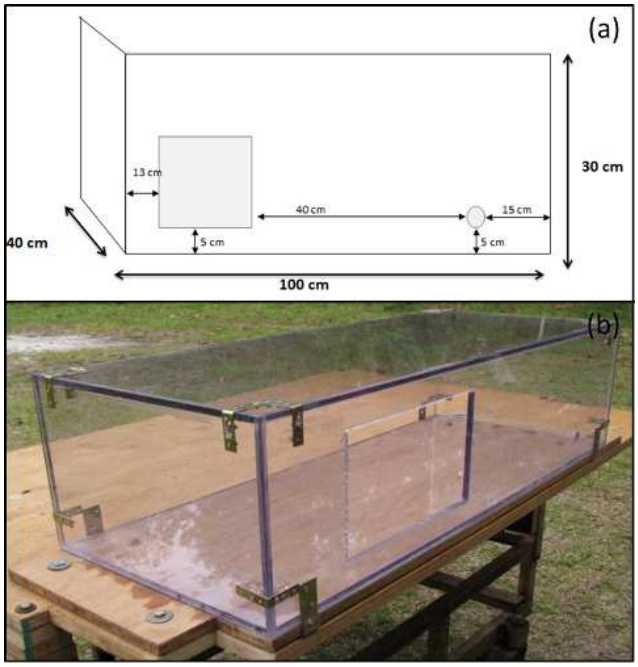
Supplementary Fig. S1: Presentation table (L 140 cm x W 100 cm x H 110 – 160 cm adjustable in height) for elevated enclosures. a) The presentation table with an additional board (L 180 cm x W 120 cm x D 3 cm) for experiments that require a vertical position. b) The table during testing of the Box Task. The individuals had access to the apparatus by reaching through the bars.

832



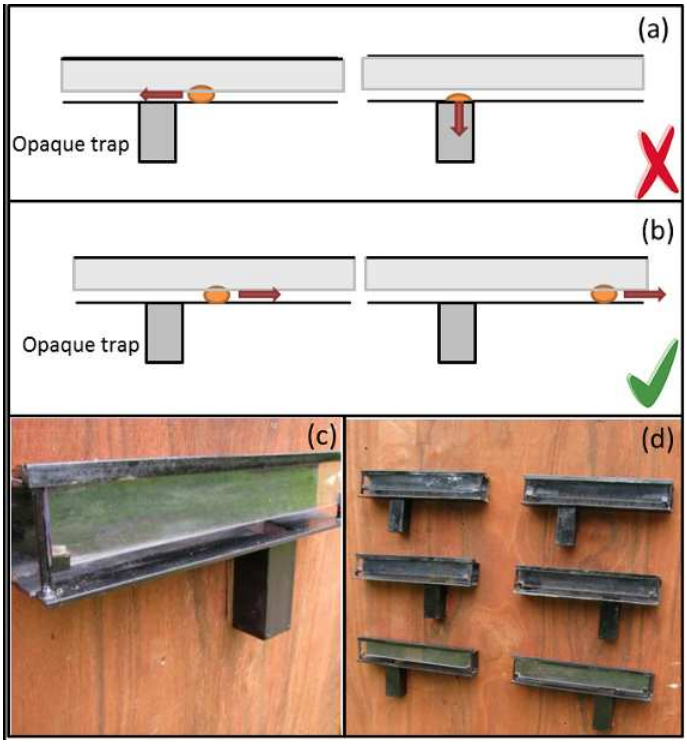
Supplementary Fig. S2: Box-Task with five wooden boxes (L 20 cm x W 15 cm x H 8 cm). (a) Schematic view of 1st presentation of the apparatus. Lid can be flipped open. (b) 2nd presentation of the apparatus. Sliding doors can be opened in order to retrieve food reward. (c) Sumatran orangutan conducting the box task, here using the sliding doors during the 2nd presentation.

833



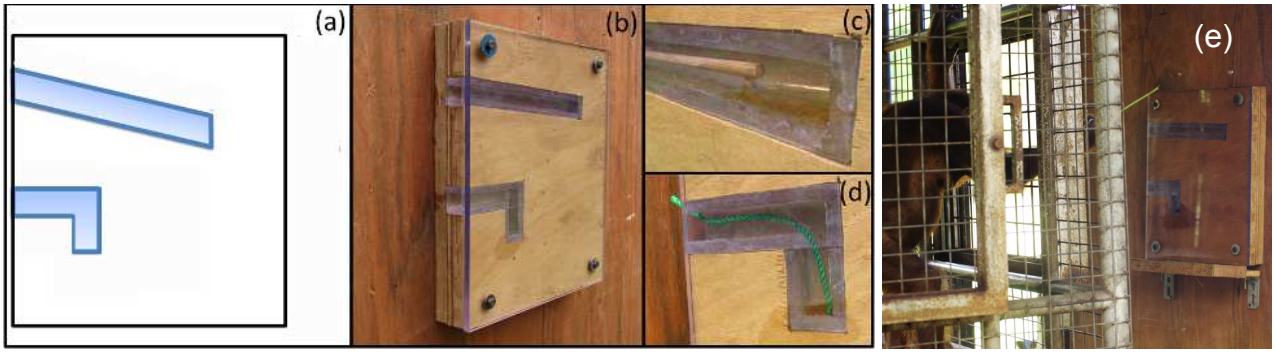
Supplementary Fig. S3: Detour Reaching. a) A schematic view including technical measurements of the apparatus used for Detour Reaching: Box (L 100 cm x W 30 cm x H 40 cm) that has two openings at the front side, one small hole (∅ 2 cm) at the right front and one large opening (L 29 cm x H 19 cm) at the left front of the box,. b) Detour Reaching Box is screwed on the presentation table and can be presented from outside of the cage. c) Sumatran orangutan solving the task by reaching for the bait, a banana, through the big hole on the left front of the box.

834



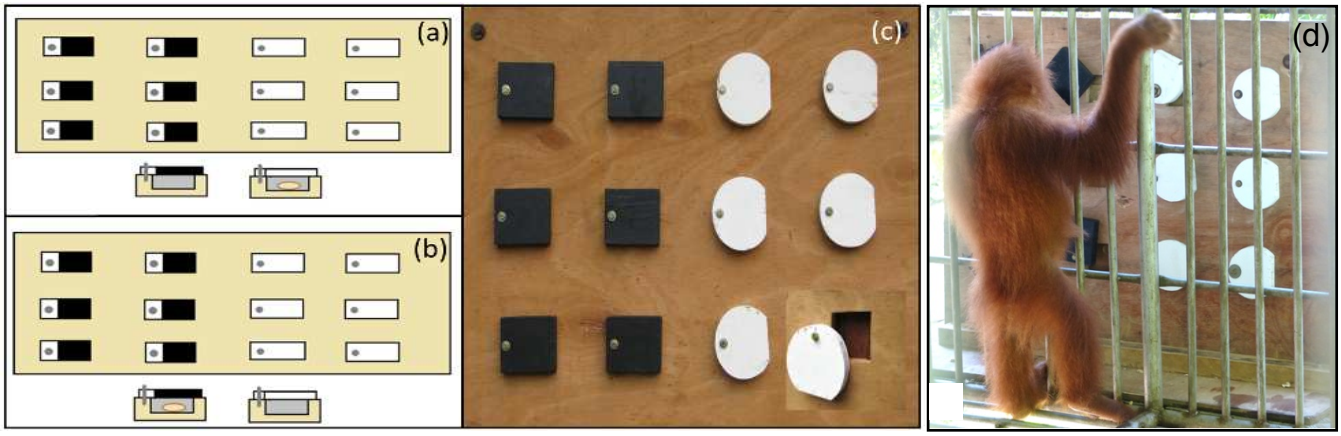
Supplementary Fig. S4: Tube Task with six tubes (L 37 x H 7.5 x D 6 cm). Subjects can use their fingers to reach trough the slot and to slide the bait along the tube. a) Wrong direction to slide the bait. Bait falls into the opaque trap. b) Correct direction to slide the bait. Bait can be retrieved out of the tube. c) Single tube, made out of metal with a transparent Makrolon plate to see the bait and run the finger in the slot. d) Complete test design, six tubes on a vertical board.

835



Supplementary Fig. S5: Visible Honey Tool Trap. a) Schematic view of the apparatus. Wooden box (H 80 cm x W 50 cm x D 7 cm) with two embedded transparent Makrolon tubes: a straight channel (L 30 cm x H 5 cm x D 5 cm) and a curved L-shaped channel [L 15 cm (10 cm until it bends down 5 cm), W 5 cm, D 5 cm]. b) Honey Tool Trap in its final version before use. c) Stick inserted in the curved channel to obtain honey (correct solution). d) Rope inserted in the curved channel in order to obtain honey (correct solution). e) Bornean individual using the stick on the apparatus.

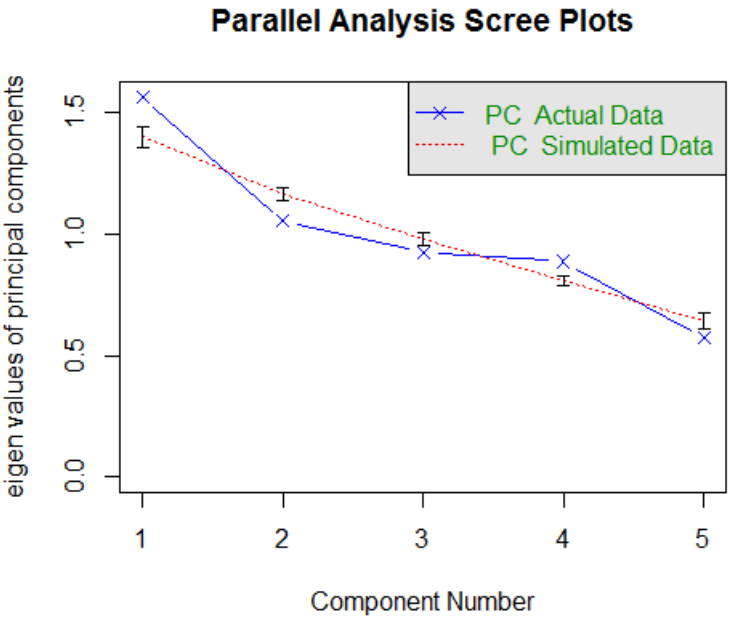
836



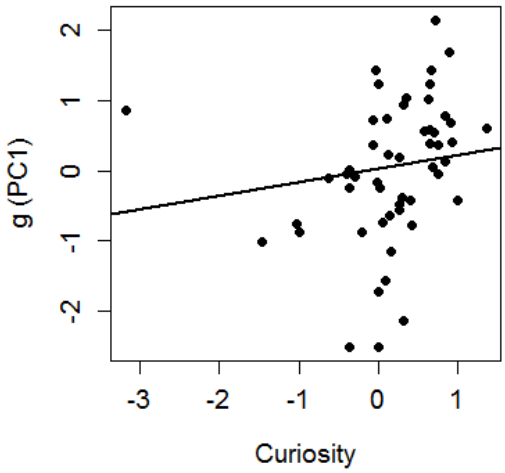
Supplementary Fig. S6: Reversal Learning with 12 doors (L 13 cm, W 10 cm, distance to each other 15 cm). (a) Schematic view of the learning phase of the Reversal Learning. The subject learns the association of food and location (here right as an example). The association with location is enhanced with colors (here white). (b) Schematic view of Reversal Phase. The old association (right, white) does not hold any more. This time the reversed association (left, black) needs to be learned. c) Reversal learning board as it is presented to the individual. d) Sumatran subject conducting the reversal learning task by reaching through the bars and turning the doors.

837

838



Supplementary Fig. S7: Parallel analysis scree plot obtained from the 'psych' package. We tested our data against randomly simulated data, estimated from 1000 randomly generated data sets with the same n and k as the empirical data set. Only one eigenvalue greater than 1 exceeded the 95% quantile and was therefore deemed statistically significant at the 95% level, yielding in one principal component.



Supplementary Fig. S8: The relation between curiosity and (PC1) plotted with a fitted line based on the LMM (Table 4) over all individuals (n=52). The outlier on the left was a highly uncurious wild individual.

839